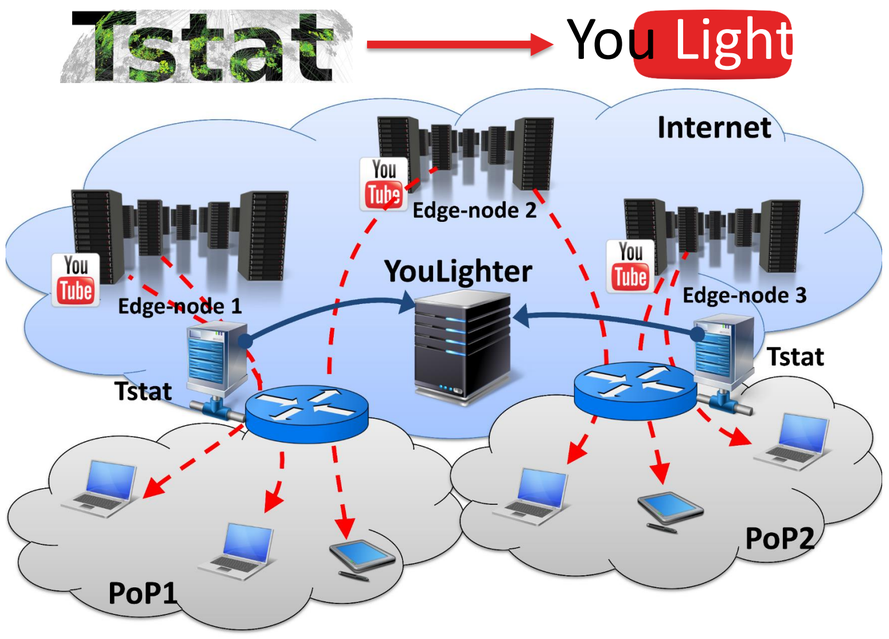
YouLighter

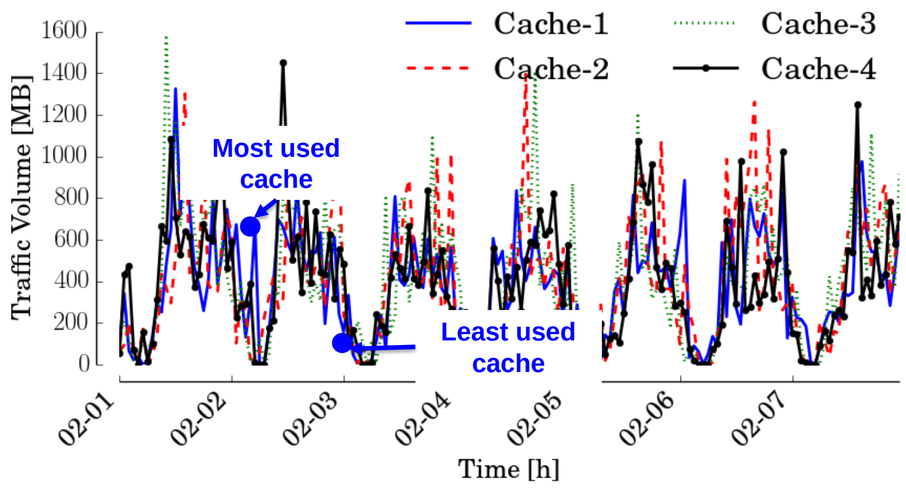
YouLighter

Paramount task of YouLighter:
- Study evolution of YouTube infrastructure
- Highlight change in YouTube infrastructure

Motivation:
- It generates 20+% of world wide traffic
- YouTube has a massive distributed infrastructure that is almost unknown
- It uses several thousands of caches (single server) grouped into Hundreds of edge-nodes
- This infrastructure suddenly evolve



Monitoring the single cache is not effective
- Load distribution changes very frequently
- The rank of most used caches **changes deeply everyday**!

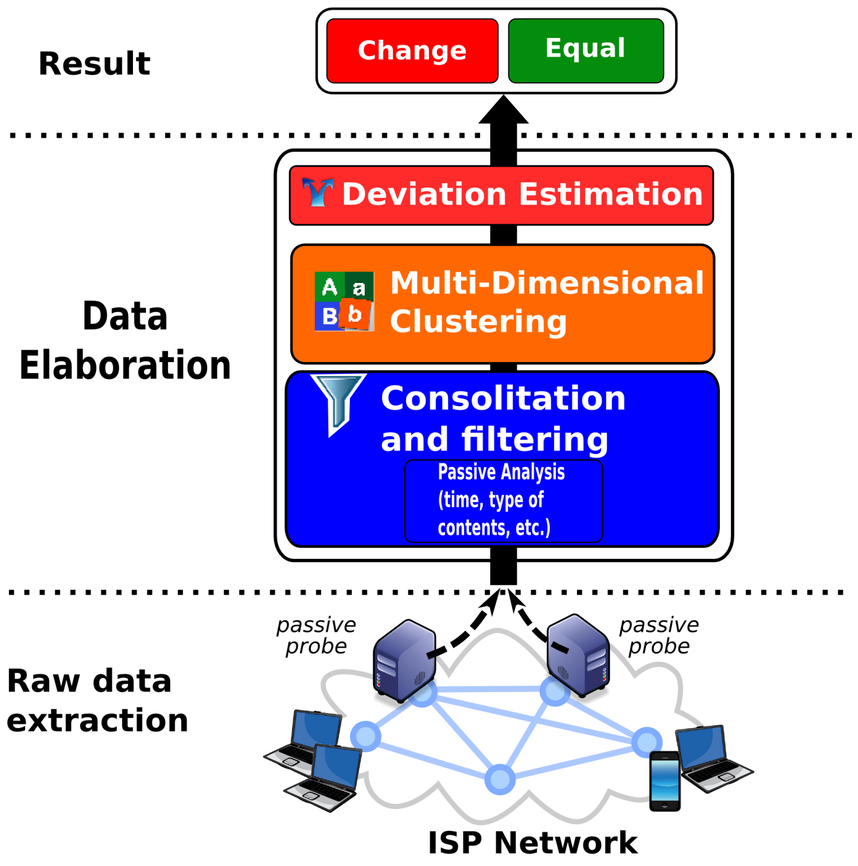


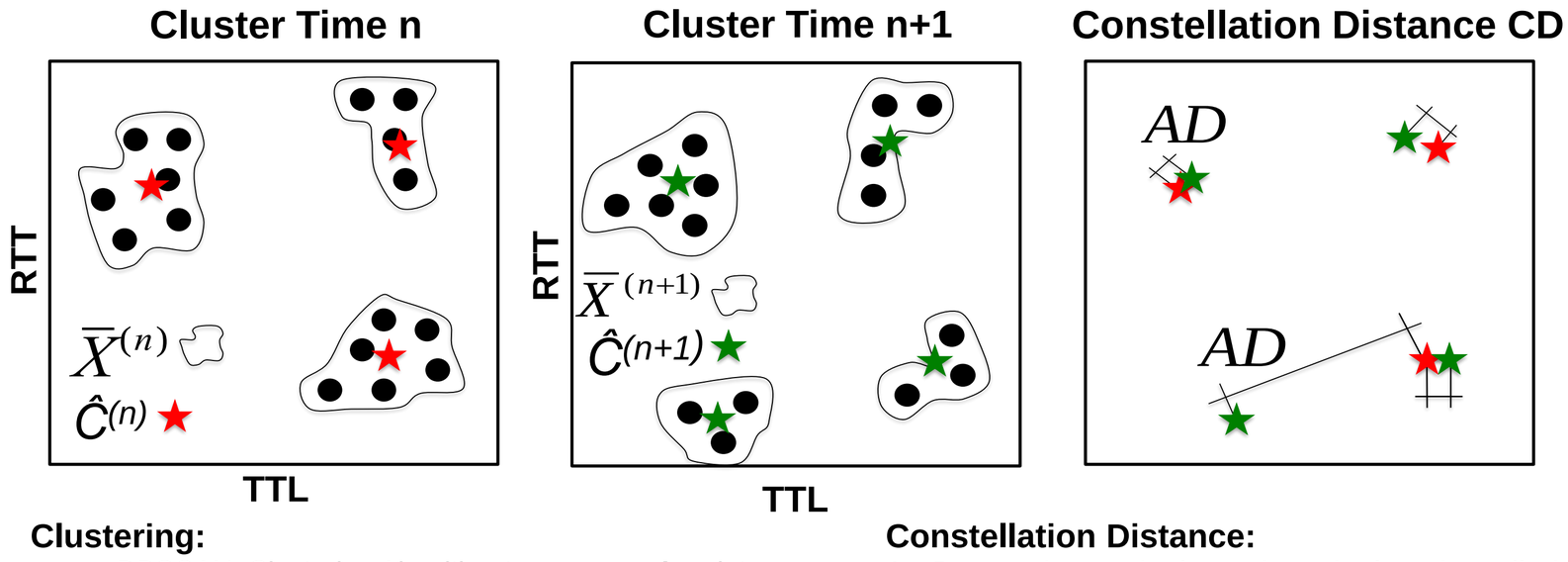
Idea: monitor **edge-nodes**, not **caches**

Dataset

Probe	Period	Volume	# Unique Videos	Caches
Probe 1 Italy	01/04/2013 - 28/02/2014	138.7 TB	2,892,452	8,664
Probe 1 Italy	01/04/2013 - 28/02/2014	152.9 TB	2,848,625	8,899
Probe 2 Italy	01/04/2013 - 28/02/2014	134.8 TB	2,711,179	9,028
Probe 3 Poland	01/03/2014 - 17/07/2014	48.3 TB	305,802	3,755

Methodology



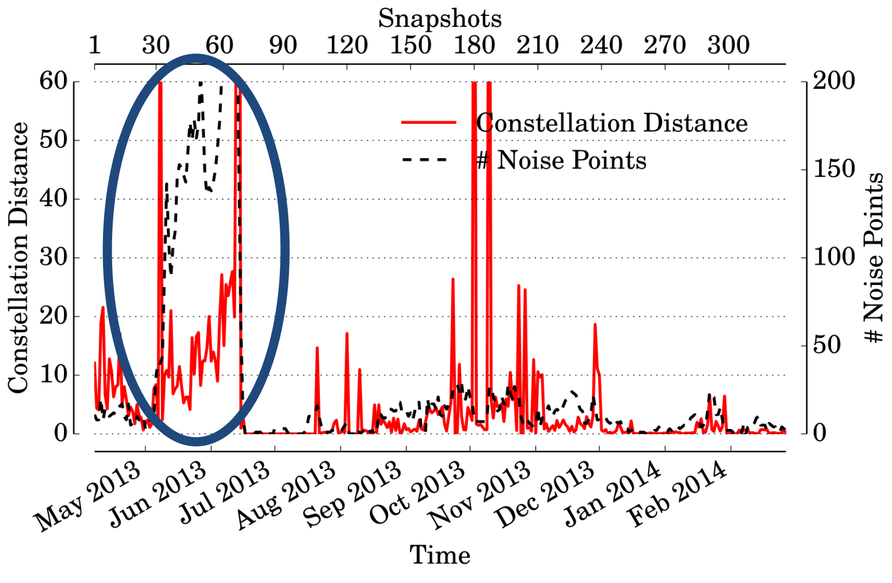


Clustering:
- DBSCAN: Clustering Algorithm to group **caches** into **adge-nodes**
- $\overline{X}^{(n)}$: Cluster result at time *n*
- $\hat{C}^{(n)}$: **Centroids** of clusters at time *n*
- *AD: Astral Distance*

Constellation Distance:
1. Summarize each cluster in a single point called **star** $\hat{C}^{(n)}$
2. ***Astra Distance***
   - For each star in $\hat{C}^{(n)}$ compute all distances to stars in $\hat{C}^{(n+1)}$ *and the* ***min***
   - *Repeat in the opposite direction*
3. ***Constellation Distance***
   - Sum all Astral Distances

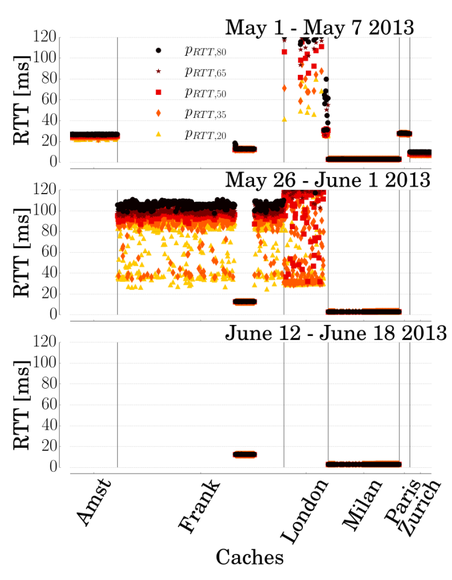
Results

Constellation Distance



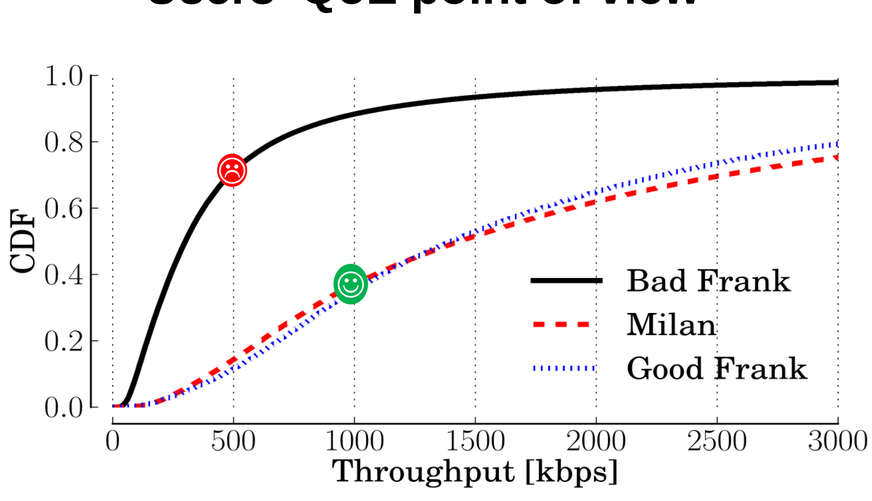
- CD Highlights either small or big
- Important changes during May – July 2013
- Changed confirmed speaking with the ISP of the probe

Network issue



- RTT before the change shows a stable path
- RTT of Frankfort during the change shows path problems for many caches
- After the change the RTT become again stable

Users' QoE point of view



- Frankfort is divided in two group based on the RTT range during the change
- Caches belonging to the group with huge range of RTT shows a Throughput distribution worst then Milan caches or caches belonging to the group with correct values of RTT

Conclusion: YouLighter shows to be effective at detecting changes in YouTube's CDN infrastructure relying on DBSCAN clustering algorithm and the novel notion of Constellation distance

Danilo Giordano, Stefano Traverso[1], Luigi Grimaudo[1], Marco Mellia[1] , Elena Baralis[1] , Alok Tongaonkar[2] and Sabyasachi Saha[2]
[1]Politecnico di Torino - first.last@polito.it, [2]Symantec - first_last@symantec.com